

日本国特許庁  
PATENT OFFICE  
JAPANESE GOVERNMENT

USPTO  
09/645404  
08/25/00

別紙添付の書類に記載されている事項は下記の出願書類に記載されて  
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed  
with this Office.

出願年月日

Date of Application: 2000年 1月18日

出願番号

Application Number: 特願2000-009165

出願人

Applicant(s): 株式会社日立製作所

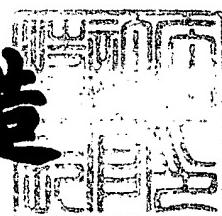
A. Imaizumi et al  
Filed 8-25-00  
703-684-1120  
ASA-915

CERTIFIED COPY OF  
PRIORITY DOCUMENT

2000年 8月11日

特許庁長官  
Commissioner,  
Patent Office

及川耕造



出証番号 出証特2000-3063718

【書類名】 特許願

【整理番号】 KN1082

【提出日】 平成12年 1月18日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 19/00

【発明者】

【住所又は居所】 愛知県尾張旭市晴丘町池上1番地 株式会社 日立製作所 情報機器事業部内

【氏名】 今泉 敦博

【発明者】

【住所又は居所】 愛知県尾張旭市晴丘町池上1番地 株式会社 日立製作所 情報機器事業部内

【氏名】 寺本 正人

【発明者】

【住所又は居所】 愛知県尾張旭市晴丘町池上1番地 株式会社 日立製作所 情報機器事業部内

【氏名】 安江 司

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社 日立製作所

【代理人】

【識別番号】 100078134

【弁理士】

【氏名又は名称】 武 頤次郎

【電話番号】 03-3591-8550

【手数料の表示】

【予納台帳番号】 006770

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1  
【物件名】 図面 1  
【物件名】 要約書 1  
【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 帳票種別判別装置

【特許請求の範囲】

【請求項1】 帳票の種別判別用のキーワード単語を予め登録した帳票種別ファイルと、

帳票のイメージデータを読み込み、読み込んだイメージデータから文字列を抽出し、抽出した文字列に対して文字認識を行う文字認識部と、

前記文字認識部が認識した文字列をキーワード単語として抽出する帳票種別判別用キーワード単語抽出部と、

前記抽出したキーワード単語と前記帳票種別ファイルに登録した帳票種別判別用のキーワード単語とを照合して、帳票の種別を判別する判別手段とからなることを特徴とする帳票種別判別装置。

【請求項2】 帳票の種別判別用のキーワード単語を予め登録した帳票種別ファイルと、

帳票のイメージデータを読み込み、読み込んだイメージデータから文字列を抽出し、抽出した文字列に対して文字認識を行う文字認識部と、

帳票の種別を表す文字列が含まれる文字列を抽出するための文字列パターンデータを格納した文字列パターンデータベースと、

前記文字認識部が認識した文字列と前記文字列パターンデータを照合して一致する単語をキーワード単語として抽出する帳票種別判別用キーワード単語抽出部と、

前記抽出したキーワード単語と前記帳票種別ファイルに登録した帳票種別判別用のキーワード単語とを照合して、帳票の種別を判別する判別手段とからなることを特徴とする帳票種別判別装置。

【請求項3】 請求項1ないし請求項2の何れか1の記載において、照合に際して抽出したキーワード単語に文字種による重み付け値を付与することを特徴とする帳票種別判別装置。

【請求項4】 請求項1ないし請求項3の何れか1の記載において、照合に際して抽出したキーワード単語に帳票上の記載位置による重み付け値を付与する

ことを特徴とする帳票種別判別装置。

【請求項5】 請求項1ないし請求項4の何れか1の記載において、照合に際して帳票の種別判別用のキーワード単語に単語種別による重み付け値を付与することを特徴とする帳票種別判別装置。

【請求項6】 請求項1ないし請求項5の何れか1の記載において、前記帳票種別ファイルに登録した帳票種別判別用のキーワード単語とを照合するキーワード単語は、抽出したキーワード単語を複数組み合わせて形成した新しいキーワード単語であることを特徴とする帳票種別判別装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は現金自動取引装置等に用いる帳票種別判別装置にかかり、特に帳票のイメージデータを読み込み、読み込んだイメージデータから文字列を抽出して帳票の種別を判別する帳票種別判別装置に関する。

【0002】

【従来の技術】

現金自動取引装置等の自動機は、自動振り込み利用申し込み書を用いた自動振り込み、口座振替票を用いた公共料金の振り替え、あるいは普通預金入金票を用いた入金処理等の各種処理を自動処理する。このとき前記自動機は利用者が挿入した自動振り込み利用申し込み書等の帳票類の種別を自動判別することが必要である。帳票種別の判別方法として最も一般的な方法は、各帳票の共通する位置に帳票種別を示すID番号、バーコード情報、マーク等の識別可能な情報を附加し、その情報を読み取ることにより帳票種別を判別する方法である。

【0003】

また、前記付加情報を必要としない判別方法としては、帳票上の特定位置にある文字列あるいはマーク等を読み取ることにより帳票種別を判別する方法、あるいは帳票上の罫線の位置、形状を読み取ることにより帳票種別を判別する方法が知られている。

【0004】

【発明が解決しようとする課題】

前記ID番号、バーコード情報、マーク等の付加情報を読み取ることにより帳票種別を判別する方法は、対象となる帳票を新規にレイアウトして作成する場合にのみ有効な方法であり、既存の帳票種別を判別する場合に適用することはできない。また、帳票上の特定の位置にある文字やマーク等を読みとことにより帳票種別を判別する方法、あるいは帳票上の罫線の位置、形状を読み取ることにより帳票種別を判別する方法は帳票レイアウトの変更やマーク形状が変更されると帳票種別を判別することができなくなる。さらに、これらの方法では帳票の印刷ずれあるいは読み取り装置のスキャン速度の変動等により画像読み取りが不安定になるおそれがある。

【0005】

本発明は前記問題点に鑑みてなされたもので、予め登録したキーワード単語と、帳票を読み取り文字認識したキーワード単語と照合することにより帳票の物理的なレイアウトの変動に対応可能な帳票種別の自動判別装置を提供する。

【0006】

【課題を解決するための手段】

本発明は、上記の課題を解決するために次のような手段を採用した。

【0007】

帳票のイメージデータを読み込み、読み込んだイメージデータから文字列を抽出し、抽出した文字列に対して文字認識を行う。次いで文字認識した文字列からなるキーワード単語と前記予め登録した帳票の種別判別用のキーワード単語とを判別手段により照合して帳票の種別を判別する。

【0008】

また、帳票のイメージデータを読み込み、読み込んだイメージデータから文字列を抽出するに際しては、前記文字認識した文字列からなるキーワード単語に対して、さらに帳票の種別を表す文字列が含まれる文字列を抽出するための文字列パターンデータを照合し、一致する単語をキーワード単語として抽出する。次いで抽出したキーワード単語と前記帳票種別ファイルに登録した帳票種別判別用のキーワード単語とを判別手段により照合して帳票の種別を判別する。

## 【0009】

## 【発明の実施の形態】

以下に本発明の実施形態を図1ないし図5を用いて説明する。図1は本実施形態にかかる帳票種別の自動判別装置の処理を説明する図である。まず、ステップ1において、帳票種別判別装置の判別対象となる帳票から抽出した帳票種別判別用のキーワード単語を登録する。

## 【0010】

図2は本発明の帳票種別判別装置の判別対象となる帳票、および該帳票から抽出した帳票種別判別用のキーワード単語を登録した帳票種別ファイルの内容を示す図である。図において、1ないし3は帳票であり、帳票1は「自動振り込み用紙（銀行控）」、帳票2は「普通預金入金票」、帳票3は「電気料金」の振り込み用紙である。また、1ないし13は帳票1ないし帳票3にそれぞれ対応した帳票種別ファイルであり、帳票1ないし3からそれぞれ帳票の種別が一意に決定できるように選択した複数のキーワード単語と、該キーワード単語のそれぞれにその重要度応じて付与された重みを登録している。

## 【0011】

例えば、帳票1は「自動振込利用申込書（銀行控）」であり、キーワード単語として「自動払込利用申込書」、「銀行控」、「口座番号」および銀行名「△〇銀行」を抽出し、抽出したキーワード単語毎にそれぞれ単語の重み値「5」、「1」、「3」および「3」を付与し、付与したファイルを帳票種別ファイル11として登録している。

## 【0012】

次に、ステップ2において、帳票種別判別装置の判別対象となる帳票1aのイメージ画像を読み取る。イメージ画像は2値化画像が望ましいが、多値化画像あるいはカラー画像であってよい。また画像読み取りに使用する光電変換手段としては、カメラやCCDセンサー等の読み取り手段を用いることができる。

## 【0013】

次に、ステップ3において、読み取ったイメージ画像中の全ての文字列を抽出する。文字列を抽出に際しては連結画素のサイズあるいは形状等の情報を利用す

ることができる。

#### 【0014】

次に、ステップ4において、イメージ画像から抽出した全ての文字列に対して文字認識を行う。

#### 【0015】

次に、ステップ5において、文字認識の結果得られた文字列中から、文字列パターンデータベース31を用いて帳票種別判別のキーワードとなるキーワード単語を抽出する。

#### 【0016】

図3は文字列パターンデータベース31の内容を示す図である。文字列パターンデータベース31には、図に示すように「\*申込書」、「\*料金」、「\*銀行」、「\*税」、「氏名」、「確認印」、「口座番号」、「銀行控」等の文字列パターンが登録されている。前記文字認識した結果と、この文字列パターンデータベースに登録されている文字列パターンデータとを照合して一致する単語をキーワード単語として抽出する。例えば、文字認識した結果である「自動振込利用申込書（銀行控）」と、文字列パターンデータベースbに登録されている文字列パターンデータ「\*申込書」とを照合すると、「自動振込利用申込書」をキーワード単語として抽出することができる。

#### 【0017】

次に、ステップ6において、前記抽出したキーワード単語に字種の重み付値および位置の重み付値を付与し、この重み付値が付与された前記キーワード単語と、ステップ1において帳票種別ファイルに登録した重み付きのキーワード単語との照合を行い、確からしさの値を求めて帳票種別の判別を行う。

#### 【0018】

このステップにおいては、まず、抽出したキーワード単語に重み付け値を付与する。図4はキーワード単語に付与する重み付け値を説明する図である。抽出したキーワード単語に対して、その字種による重み付け値を付与する。キーワード単語の字種は、文字列の直線性や文字の間隔などの特徴を検出して活字であるか手書きであるかを判別することにより行い、判別した字種にしたがってキーワー

ド単語の重み付けを行う。本実施形態では、帳票種類判別のために活字のみを使用し手書き文字は使用しないこととするので、活字文字であれば重み付け値を1とし、手書き文字であれば重み付け値を0とする。

#### 【0019】

また、抽出したキーワード単語が帳票内で記載されていた位置に応じて重み付けを行う。本実施形態では、図4に示すように帳票を縦方向に等間隔に10の領域に分割し、帳票の上部に記載された文字列ほどその帳票を特徴付ける文字列が多いものとして、最上部の領域には10の重み付け値を付与し、以下、キーワード単語の記載領域にしたがって9から1までの位置の重み付け値を付与する。

#### 【0020】

次に、帳票種別の判別を行う。帳票種別の判別に際しては、前記字種の重み付値および位置の重み付値が付与された前記キーワード単語と、帳票種別ファイルに登録した重み付きのキーワード単語との照合を行い、確からしさの値を求めて帳票種別の判別を行う。

#### 【0021】

本実施形態では、以下の計算式を用いて、帳票の確からしさの値を求める。

#### 【0022】

K = 抽出キーワード単語の文字種別による重み

P = 抽出キーワード単語の記載位置による重み

J = 帳票種別ファイルに登録されている重み

確からしさの値 = K × P × J

帳票の確からしさの値の計算は、照合する全ての単語について前記計算式により確からしさの値求め、求めた値の合計を帳票の確からしさの値とし、確からしさの値が最も高い帳票を入力画像の帳票種別として判別する。

#### 【0023】

図5は帳票の確からしさの値の計算例を示す図である。図において、帳票1、帳票2、帳票3の確からしさの値がそれぞれ72、9、12となり、帳票1の72が最も大きな値となっている。したがって、入力画像の帳票種類を帳票1と判別する。

**【0024】**

なお、以上の説明では、ステップ1において記帳票種別ファイルに登録した帳票種別判別用のキーワード単語と、ステップ5において抽出したキーワード単語とを照合したが、ステップ5において抽出したキーワード単語に代えて、ステップ5において抽出したキーワード単語を複数組み合わせて生成した新しいキーワード単語としてもよい。

**【0025】**

図6は単語を組み合わせて新しいキーワード単語を形成する手順を説明する図である。図において1 bは判別対象となる帳票、11 aは帳票種別ファイルである。新しいキーワード単語の生成に際しては、まず、判別対象となる帳票1 bからキーワード単語「平成〇〇年」、「納税通知書」・・・「〇×市」、「市長」を抽出する。次いで抽出したキーワード単語「平成〇〇年」、「納税通知書」・・・「〇×市」、「市長」を組み合わせて、例えば「〇×市納税通知書」を新しいキーワード単語として生成する。次いでこの新しいキーワード単語を前記記帳票種別ファイル11 aに登録した帳票種別判別用のキーワード単語と照合して帳票1 bの帳票種別を判別する。

**【0026】**

以上のように、本実施形態によれば、帳票種別を判別するために任意の位置にある単語を自動抽出して文字認識し、帳票種別情報に記載された単語群と照合することで帳票識別を行うため、下記の効果を奏すことができる。

**【0027】**

1) 帳票に対してバーコード、ID番号等の新たな情報を付加しなくとも、帳票種別を判別することが可能となる。

**【0028】**

2) 帳票レイアウトの変更、帳票のフォントの変更があっても、帳票種別を判別することが可能となる。

**【0029】**

3) 帳票の印刷ずれがあっても、帳票種別を判別することが可能となる。

**【0030】**

4) 帳票判別のための帳票の特徴情報の登録を容易に行うことができる。また、帳票判別時の特徴情報を記憶する格納領域を低減することができる。

【0031】

5) 帳票内の任意の位置の文字列を用いるため、帳票種別判別を行う自由度が高くなり、同時に判別可能な帳票の種別を増加させることができる。

【0032】

【発明の効果】

以上説明したように本発明によれば、予め登録したキーワード単語と、帳票を読み取り文字認識したキーワード単語と照合するので、帳票の物理的なレイアウトの変動に対応可能な帳票種別の自動判別装置を提供することができる。

【図面の簡単な説明】

【図1】

本発明の実施形態にかかる帳票種別の自動判別装置の処理を説明する図である。

【図2】

帳票種別ファイルの内容を説明する図である。

【図3】

文字パターンデータベースの内容を示す図である。

【図4】

キーワード単語に付与する重み付け値を説明する図である。

【図5】

帳票の確からしさの値の計算例を示す図である。

【図6】

単語を組み合わせて新しいキーワード単語を形成する手順を説明する図である。

【符号の説明】

1, 2, 3, 1a, 1b 帳票

特2000-009165

11, 12, 13, 11a 帳票種別ファイル

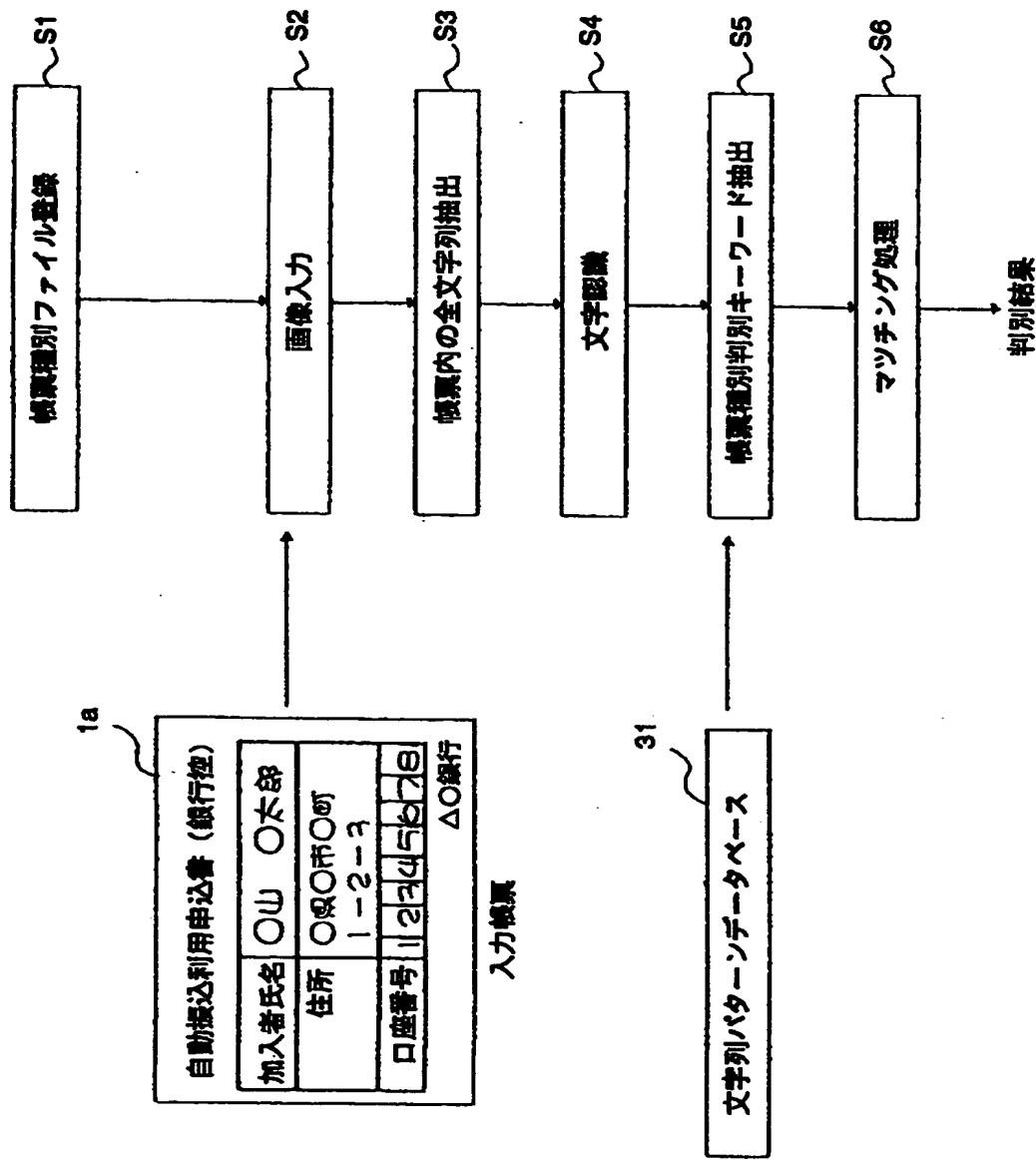
3, 3a 文字列パターンデータベース

【書類名】

図面

【図1】

図1



【図2】

図2

帳票1

自動振込利用申込書（銀行控）

加入者氏名							
住所							
口座番号	1	2	3	4	5	6	7
△○銀行							

1



帳票1

キーワード単語	単語の重み
自動振込利用申込書	5
銀行控	1
口座番号	3
△○銀行	3

11

12

帳票2

普通預金入金票

加入者氏名							
金額							
口座番号	1	2	3	4	5	6	7
○×銀行							

2



帳票2

キーワード単語	単語の重み
普通預金入金票	5
○×銀行	3
口座番号	3
金額	1

13

帳票3

電気料金 △○銀行 ××支店

口座番号	1	2	3	4	5	6	7
電気料金							
加入者氏名							

3



帳票3

キーワード単語	単語の重み
電気料金	5
口座番号	3
△○銀行	3

【図3】

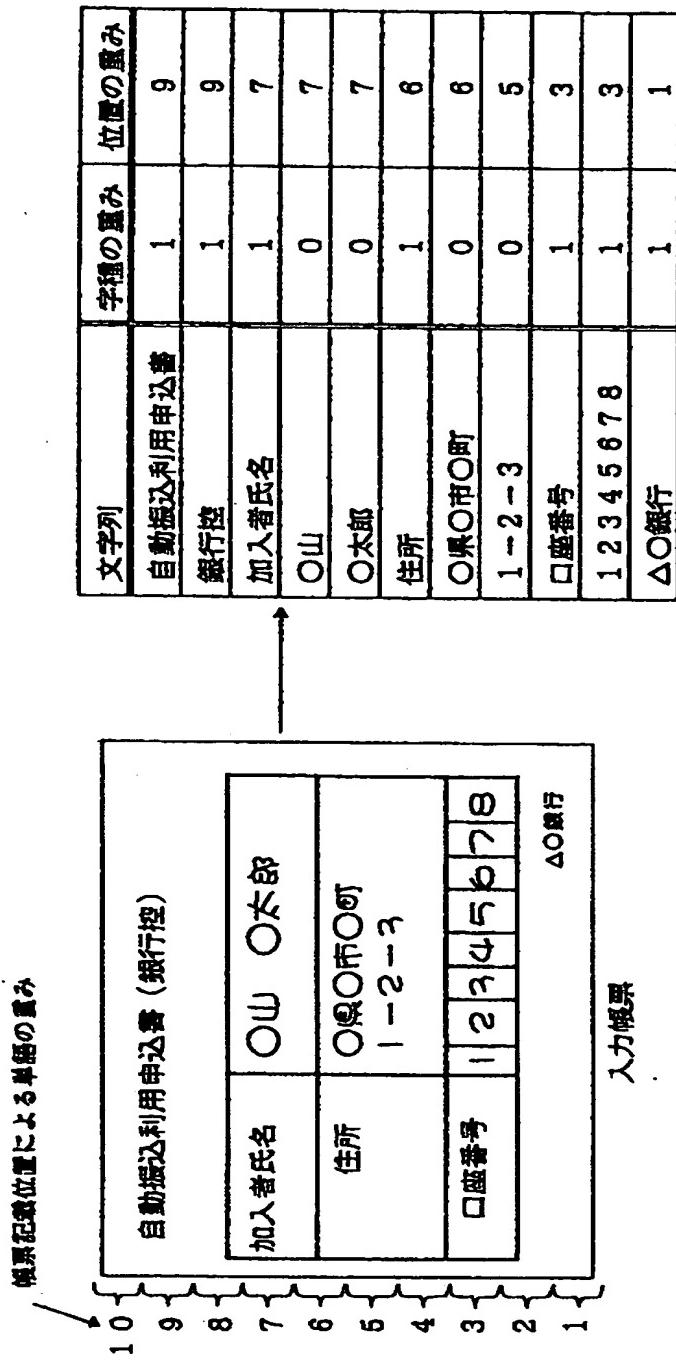
図3

31

*申込書
*料金
*銀行
*税
氏名
確認印
口座番号
銀行控
:
:

【図4】

図4



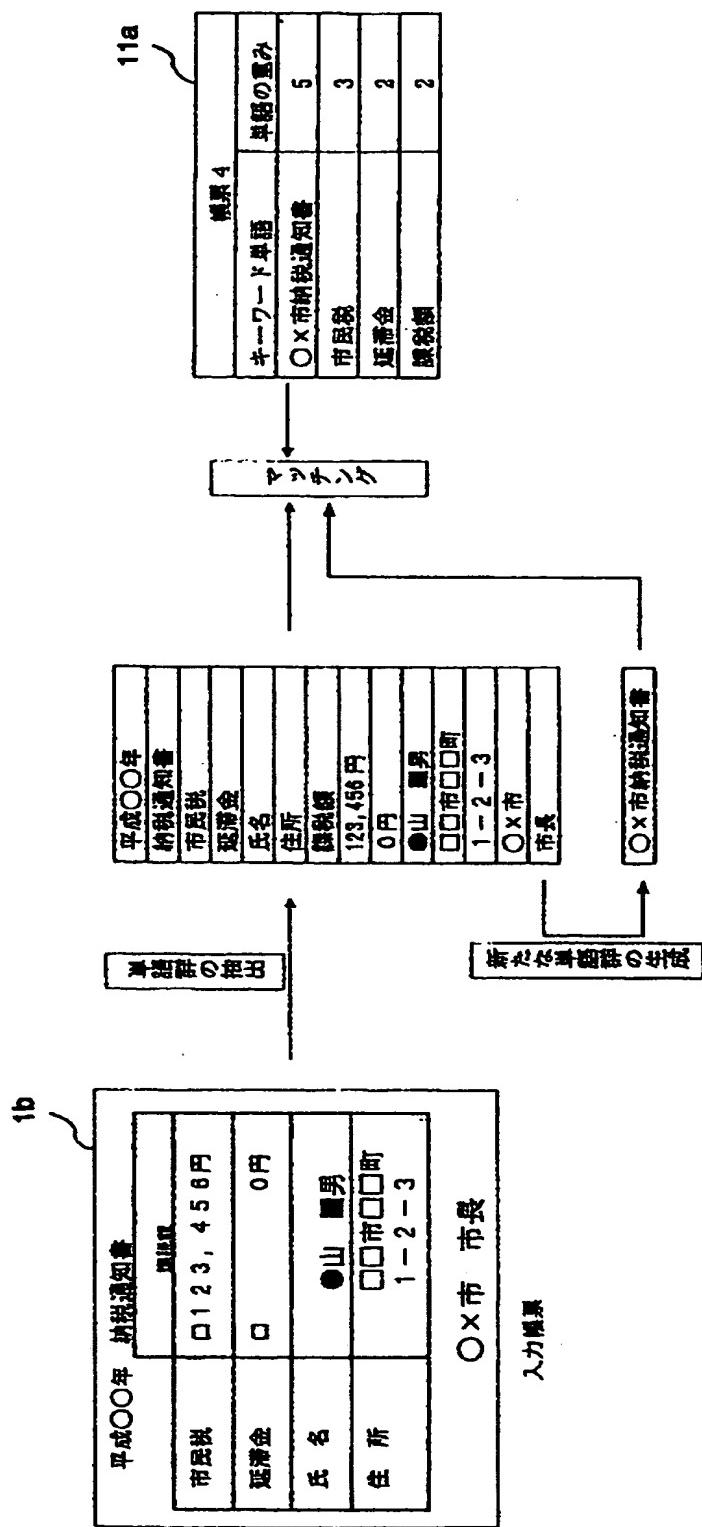
【図5】

図5

文字列	字種の重み	位置の重み	帳票1 単語の重み	帳票2 単語の重み	帳票3 単語の重み
自動振込利用申込書	1	9	5		
銀行控	1	9	1		
加入者氏名	1	7			
○山	0	7			
○太郎	0	7			
住所	1	6			
○県○市○町	0	6			
1-2-3	0	5			
口座番号	1	3	3	3	3
12345678	0	3			
△○銀行	1	1	3		3
盤からしさの値			72	9	12

【図6】

図6



【書類名】 要約書

【要約】

【課題】 予め登録したキーワード単語と、帳票を読み取り文字認識したキーワード単語と照合することにより、帳票の物理的なレイアウトの変動に対応可能な帳票種別の自動判別装置を得る。

【解決手段】 帳票の種別判別用のキーワード単語を予め登録した帳票種別ファイル1と、帳票のイメージデータを読み込み、読み込んだイメージデータから文字列を抽出し、抽出した文字列に対して文字認識を行う文字認識部4と、前記文字認識部が認識した文字列をキーワード単語として抽出する帳票種別判別用キーワード単語抽出部5と、前記抽出したキーワード単語と前記帳票種別ファイルに登録した帳票種別判別用のキーワード単語とを照合して、帳票の種別を判別する判別手段6とからなる。

【選択図】 図1

出願人履歴情報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日

[変更理由] 新規登録

住 所 東京都千代田区神田駿河台4丁目6番地

氏 名 株式会社日立製作所